

卡方分布

维基百科，自由的百科全书

卡方分布（**chi-square distribution**, **χ²-distribution**）是概率论与统计学中常用的一种概率分布。**k**个独立的标准正态分布变量的平方和服从**自由度**为**k**的卡方分布。**卡方分布**是一种特殊的伽玛分布。假设检验和置信区间的计算。

由卡方分布延伸出来皮尔森卡方检定常用于：
(1)样本某性质的比例分布与总体理论分布的拟合优度；
(2)同一总体的两个随机变量是否独立；
(3)二或多个总体同一属性的同素性检定。（详见皮尔森卡方检定）

目录

▪ 1 数学定义

▪ 2 特征

▪ 3 Table of χ² value vs p-value

▪ 4 参考文献

▪ 5 外部链接

数学定义

若**k**个随机变量**Z₁**、.....、**Z_k**是相互独立，符合标准正态分布的随机变量（数学期望为0、方差为1），则随机变量**Z**的平方和

$$X=\sum_{i=1}^kZ_i^2$$

被称为服从自由度为 **k** 的卡方分布，记作

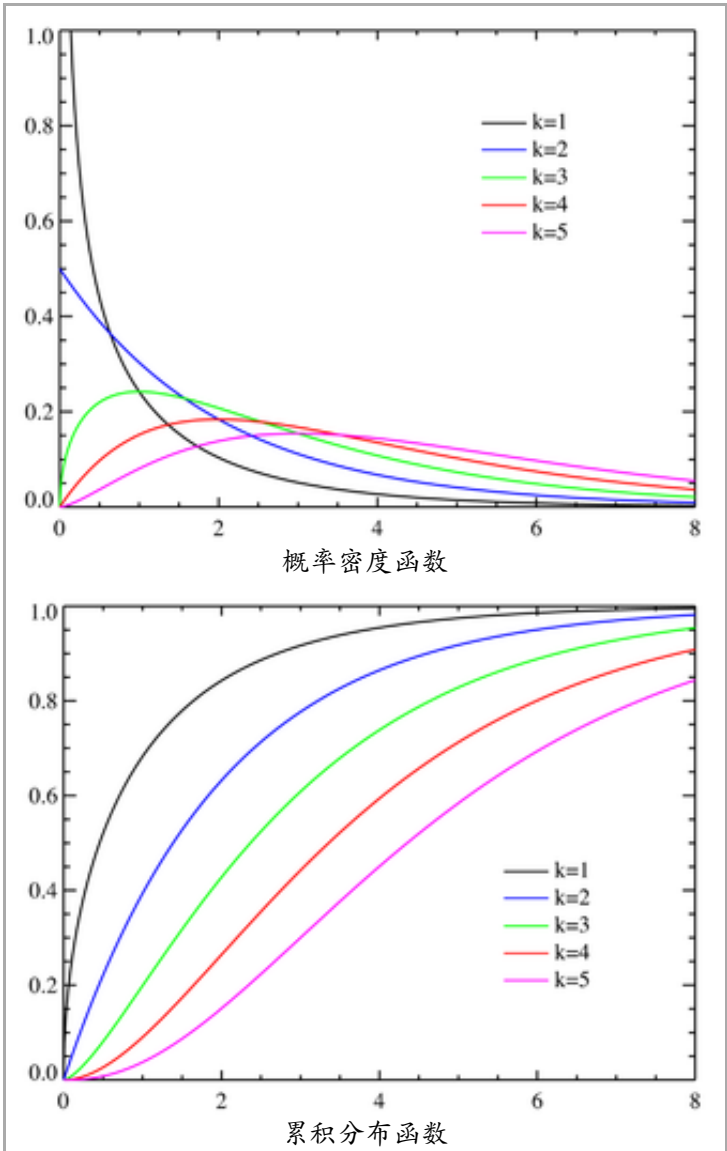
$$\begin{aligned} X &\sim \chi^2(k) \\ X &\sim \chi_k^2 \end{aligned}$$

特征

卡方分布的概率密度函数为：

$$f_k(x)=\frac{(1/2)^{k/2}}{\Gamma(k/2)}x^{k/2-1}e^{-x/2}$$

卡方分布



参数	$k > 0$ 自由度
支撑集	$x \in [0; +\infty)$,
概率密度函数	$\frac{(1/2)^{k/2}}{\Gamma(k/2)}x^{k/2-1}e^{-x/2}$,
累积分布函数	$\frac{\gamma(k/2, x/2)}{\Gamma(k/2)}$,
期望值	k ,
中位数	大约 $k - 2/3$,
众数	$k - 2$, 如果 $k \geq 2$,
方差	$2k$,
偏度	$\sqrt{8/k}$,
峰度	$12/k$,
信息熵	

其中 $x \geq 0$ ，当 $x \leq 0$ 时 $f_k(x) = 0$ 。这里 Γ 代表Gamma函数。

卡方分布的累积分布函数为：

$$F_k(x) = \frac{\gamma(k/2, x/2)}{\Gamma(k/2)},$$

其中 $\gamma(k, z)$ 为不完全Gamma函数

在大多数涉及卡方分布的书中都会提供它的累积分布函数的对照表。此外许多表格计算软件如OpenOffice.org Calc和Microsoft Excel中都包括卡方分布函数。

自由度为 k 的卡方变量的平均值是 k ，方差是 $2k$ 。卡方分布是伽玛分布的一个特例，它的熵为：

$$H = \int_{-\infty}^{\infty} f(x) \ln(f(x)) dx = \frac{k}{2} + \ln \left(2\Gamma \left(\frac{k}{2} \right) \right) + \left(1 - \frac{k}{2} \right) \psi(k/2)$$

其中 $\psi(x)$ 是双伽玛函数。

卡方变量与Gamma变量的关系 当Gamma变量 频率（ λ ）为1/2时， α 的2倍为卡方变量之自由度（Degree of freedom） 即：

$$\begin{aligned} r.v.Y = \chi^2(U) &= \Gamma \left(\frac{U}{2}, \frac{1}{2} \right) \\ E \left(\chi^2(U) \right) &= E(Y) = \frac{\alpha}{\lambda} = \frac{\frac{U}{2}}{\frac{1}{2}} = U \\ Var \left(\chi^2(U) \right) &= Var(Y) = \frac{\alpha}{\lambda^2} = \frac{\frac{U}{2}}{\left(\frac{1}{2}\right)^2} = 2U \end{aligned}$$

卡方变量之期望值=自由度 卡方变量之方差=两倍自由度

Table of χ^2 value vs p-value

p-value = 1- p_CDF.

χ^2 越大: p-value越小, 可信度越高. 通常用 $p=0.05$ 作为threshold, 即95%的可信度. 常用的 χ^2 与p-value表如下:

$$\frac{k}{2} + \ln(2\Gamma(k/2)) + (1 - k/2)\psi(k/2)$$

动差生成函数

$$(1 - 2t)^{-k/2}, 2t < 1,$$

特性函数

$$(1 - 2it)^{-k/2},$$

数

自由度	<i>χ</i> ² value ^[1]											
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83	
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82	
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27	
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47	
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52	
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46	
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32	
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12	
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88	
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59	
P value（概率）	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001	

参考文献

1. ^ Chi-Squared Test (<http://www2.lv.psu.edu/jxm57/irp/chisquar.html>) Table B.2. Dr. Jacqueline S. McLaughlin at The Pennsylvania State University. In turn citing: R.A. Fisher and F. Yates, Statistical Tables for Biological Agricultural and Medical Research, 6th ed., Table IV

外部链接

- 分布计算器 (http://www.vias.org/simulations/simusoft_distcalc.html)（英文）

取自“<http://zh.wikipedia.org/w/index.php?title=卡方分佈&oldid=32577942>”

■ 本页面最后修订于2014年9月6日 (星期六) 23:21。

■ 本站的全部文字在知识共享 署名-相同方式共享 3.0协议之条款下提供，附加条款亦可能应用（请参阅使用条款）。

Wikipedia®和维基百科标志是维基媒体基金会的注册商标；维基™是维基媒体基金会的商标。维基媒体基金会是在美国佛罗里达州登记的501(c)(3)免税、非营利、慈善机构。